A preconditioner based on a low-rank approximation with applications to topology optimization

Paolo Gatto, Rasmus E. Christiansen, Jan S. Hesthaven

April 9, 2015

Abstract

Probabilistic algorithms for constructing matrix decompositions have recently gained vast popularity due to their ability to handle very large problems. This paper pursues the idea of improving a standard Jacobi preconditioner with a low rank correction, to enhance its performance for the iterative solution of finite element discretizations. This is based on the well-known fact that, in the case of elliptic problems, the Green function is numerically low-rank. We follow the same approach in the context of wave propagation problems. Although the Green function is no longer low-rank, we show that, in the case of moderate wave numbers, a low-rank update can be successfully employed to construct an effective preconditioner. We investigate the behavior of such preconditioner when applied to subsequent discretizations of the Helmholtz equation that arise in the context of topology optimization.

Keywords. Green function, Interpolative Decomposition, topology optimization.

1 Introduction

Classical linear algebra techniques are ill-suited for applications that yield extremely large matrices whose entries are inherently affected by errors. Consequently, the employment of a highly accurate matrix decomposition makes little sense when the efficient handling of large amounts of data becomes paramount. This new scenario calls for randomized techniques and provides the opportunity to design algorithms that fit recent developments in computer
hardware, where the role of data transfer is central to achieve efficiency. For an extensive survey of randomized matrix approximations see [11].

Among randomized matrix decompositions, the Interpolative Decomposition (ID) is a powerful tool that allows to compress a numerically low-rank matrix $A$. In fact, for every $\varepsilon$, it is possible to determine a skeletonization $A^{(skel)}$, i.e., a collection of $k$ columns of $A$, such that $\|A - A^{(skel)} P\| \leq \varepsilon$, where $P$ contains the $k$-order identity as a submatrix. In the case of well-behaved elliptic problems posed on a domain $\Omega$, the approximation $G(x_i, y_j)$ to the Green function, where $\{x_i\}, \{y_j\}$ are sets of points in $\Omega$, is numerically low-rank. This property is commonly exploited in the context of boundary integral methods, in which a boundary value problem (BVP) is recast as a boundary integral equation (BIE) involving a Fredholm operator of the second kind. If $A$ is the matrix approximation of the integral in the BIE, then the matrix-vector product $Av$ can be evaluated cheaply. As a result, Fast Multipole Methods of linear complexity have been developed for a variety of kernels (heat equation, wave equation, Stokes equations, etc.), see [13, 14, 9].

In this paper we apply similar ideas to the context of Finite Elements (FE) discretizations. When $A$ is the stiffness matrix arising from a FE discretization of an elliptic problem, the off-diagonal part of $A^{-1}$ is a low-rank matrix as a result of the decay of the Green function. We exploit this property to build a preconditioner by adding a low-rank update to a banded Jacobi preconditioner. As we transition from elliptic to hyperbolic problems, more specifically to the Helmholtz equation, the behavior of the Green function changes such that a low rank update, a priori, no longer exists or, rather, $k$ is no longer guaranteed to be small when compared to the size of the matrix. Nevertheless, if we apply this same machinery in the case of low frequencies $\omega$ or, equivalently, low wave numbers $\kappa$, numerical evidence demonstrates that, even for modest values of $k$, the preconditioner remains effective.

Linear systems that arise from the discretization of wave propagation phenomena are typically highly indefinite, which makes the construction of an effective preconditioner both vital and challenging. Standard multigrid methods generally do not work well, mainly because of the fact that the oscillations on the scale of the wavelength cannot be carried on to the coarse grids. Incomplete $LU$ decompositions are fairly expensive to compute and still lead to a number of iterations that scales linearly with respect to $\omega$. Recently, a class of methods that rely on the idea of preconditioning via a shifted Laplacian has gained popularity, see [8, 7, 12]. Although those methods lead to significant improvements, the iteration count still grows linearly with $\omega$. The latest development is the so-called “sweeping preconditioner” of Engquist and Ying, see [6, 5], developed in the context of finite difference approximations. It employs the compressibility of the half-space Green function to produce
an approximate factorization, sweeping the domain layer by layer. Both the construction and the application of the preconditioner are of linear complexity.

The construction of the sweeping preconditioner rests upon a specific ordering of the degrees of freedom, namely layer by layer, hence the name, and exploits the physical meaning of the tridiagonal block structure of the resulting matrix. The goal of this work is to develop a preconditioner that is still robust when such a tight connection between physics and numerical approximation cannot be established. In fact, the only core assumption of our construction is that a “reasonable” ordering of the degrees of freedom has been selected, e.g., by means of a space-filling curve, so that, to a set tolerance $\varepsilon$, the ID can be determined with an effectively small numerical rank $k$.

The paper is organized as follows. In Section 2 we review the Interpolative Decomposition and some facts from linear algebra. In Section 3 we discuss the construction of the preconditioner in details and give a few illustrative numerical examples. Section 4 is dedicated to an application of the preconditioner to a topology optimization problem. Finally, in Section 5 we draw conclusions from the presented work and point to possible future directions of research.

2 Facts from linear algebra

In numerical linear algebra, producing a low-rank approximation to a matrix is a central question. To make the discussion precise, for a positive error tolerance $\varepsilon$, we wish to construct an approximation $X$ to a given matrix $A$, such that $\|A - X\| \leq \varepsilon$ and rank$(X)$ is as small as possible. Henceforth, $\|\cdot\|$ indicates the spectral norm. A natural answer is provided by the Singular Value Decomposition (SVD): if $k$ is the number of singular values of $A$ that exceed $\varepsilon$, the choice $X = QQ^*A$, where the columns of $Q$ are the $k$ dominant left singular vectors of $A$, yields the minimizer of $\|A - X\|$ over all matrices of rank $k$.

A different way to realize a rank-$k$ approximation of $A$ is through an Interpolative Decomposition (ID). In simple words, for every $\varepsilon > 0$, it is possible to determine a collection of $k$ columns of $A$, i.e., a “skeletonization” $A^{(skel)}$, such that $\|A - A^{(skel)}P\| \leq \varepsilon$, where some subset of $k$ columns of $P$ makes up the identity matrix and no entry of $P$ has absolute value greater than 1. The decomposition is interpolative in the sense that each column of $A$ can be expressed as a linear combination of $k$ fixed columns with bounded coefficients.

The precise statement is provided by the following theorem, see [19], Lemma 3.1.

**Lemma 2.1.** For every $m \times n$ complex matrix $A$ and every integer $k$ such
that \( k \leq \min\{m, n\} \), there exist a complex \( k \times n \) matrix \( P \) and a complex \( m \times k \) matrix \( B \), whose columns constitute a subset of the columns of \( A \), such that:

1. some subset of the columns of \( P \) makes up the identity of order \( k \);
2. no entry of \( P \) has an absolute value greater than 1;
3. \( \|P\| \leq \sqrt{k(n-k)+1} \);
4. the least (that is the \( k \)-th greatest) singular value of \( P \) is at least 1;
5. \( BP = A \) when \( k = m \) or \( k = n \);
6. \( \|BP - A\| \leq \sqrt{k(n-k)+1} \sigma_{k+1} \) when \( k < m \) and \( k < n \), where \( \sigma_{k+1} \) is the \((k+1)\)-st greatest singular value of \( A \).

Existing algorithms for computing \( B \) and \( P \) that satisfy the properties listed above are computationally expensive. In [19], an algorithm whose cost is proportional to \( mn \log(k) + l^2(m+n) \) floating point operations (here \( l \) is an integer near to, but greater than, \( k \)), is obtained by slightly weakening the conditions on the factors \( B \) and \( P \):

2′. no entry of \( P \) has an absolute value greater than 2;
3′. \( \|P\| \leq \sqrt{4k(n-k)+1} \);
6′. \( \|BP - A\| \leq \sqrt{4k(n-k)+1} \sigma_{k+1} \).

Furthermore, the algorithm parallelizes trivially.

### 3 Preconditioner

The Krylov methods are projection methods that minimize the residual norm over the affine subspace \( x_0 + \mathcal{K}_m \), where \( x_0 \) is an initial guess for the solution, and \( \mathcal{K}_m \) is the \( m \)-th Krylov subspace, see [15]. Some variants such as the GMRES method are guaranteed to converge in a number of iterations equal to the size of the system. However, its restarted version, which is used in practice, tends to stagnate in the case of an indefinite matrix. This shortcoming is usually overcome through the use of an appropriate preconditioner. In this work we shall employ the left-preconditioned GMRES.

A Jacobi preconditioner is perhaps the simplest of all preconditioners. In the standard splitting \( A = M - N \), see, e.g., [15], it corresponds to the choice \( M = \text{diag}(A) \). A simple variant is a banded Jacobi preconditioner, which is built from the band of \( A \), namely the diagonal and \( b^1 \) sub- and super-diagonals. While extremely cheap to apply, a Jacobi preconditioner is effective only in a limited number of cases, e.g., diagonally dominant, symmetric positive definite matrices.

---

1Here \( b \) is an integer called bandwidth. In the case \( b = 0 \), a standard Jacobi preconditioner is recovered.
Following the approach outlined in [2], we build our preconditioner as a low-rank correction to a banded Jacobi preconditioner $G$. The low-rank correction $\delta G$ is obtained from a skeletonization of $A$, i.e., a collection of its columns.

The construction is as follows. For some bandwidth $b$, let $G$ be the inverse of the band of $A$. We proceed to construct a matrix $Y$ whose range is a sample of the range of $A^{-1} - G$. Let $l$ be a sampling parameter and $N$ be an $l \times m$ random matrix whose entries are drawn from independent trials of a normal distribution with zero mean and unit variance. We build $Y$ by applying the random matrix $R := NA$, as:

\[ Y := R(A^{-1} - G) = N - NAG \]

As a next step, we perform an ID of $Y$ to a desired tolerance $\varepsilon$:

\[ \|Y - Y^{(skel)}P\| \leq \varepsilon \]

where $Y^{(skel)}$ is an $m \times k$ matrix comprising $k$ columns of $Y$. As a final and most critical step, we construct an $m \times k$ matrix $B$ by selecting $k$ columns of $A^{-1} - G$ in the same fashion as the columns of $Y^{(skel)}$ were chosen from those of $Y$. The construction of $B = [u_1 | \cdots | u_k]$ is tantamount to the solution of linear systems:

\[ Au_i = (I - AG)(:,j_i) \quad , \quad i = 1, \ldots, k \]

where \( \{j_1, \ldots, j_k\} \) are the indices of the columns of $Y$ that form $Y^{(skel)}$. Although the construction of $\delta G$ is $k$ times more expensive than a single linear solve, this is a trivially parallelizable process. Finally, we formally set $\delta G = BP$, although such matrix is never formed explicitly. The whole procedure is detailed in Algorithm 1.

As a first test case, we investigate the performance of the preconditioner on a finite element (FE) approximation of the Poisson problem \(-\Delta u = f\) posed on the unit square $\Omega = (0, 1)^2$. Rather than solving a specific problem, we are interested in the behavior of the preconditioner with respect to the discretization of the differential operator. Thus, we set the load vector equal to a vector with unit entries. We employ uniform $n \times n$ quadrilateral grids with constant order of approximation $p$ and elements ordered lexicographically. As is customary, within each element we enumerate the modes from high to low order. Figure 1 shows typical sparsity patterns of the resulting stiffness matrices. The choice of the banded Jacobi preconditioner suggests itself, namely $b = (p + 1)^2 - 1$. As a consequence of the element ordering, this is an $n$-block diagonal matrix. We

\[ \text{This is to say that, if each element is identified to its centroid } (\frac{i}{n}, \frac{j}{n}), \text{ the ordering is } (j - 1)n + i. \]
Algorithm 1: construction of $\delta G$.

Input: $l, \varepsilon$

build an $l \times m$ standard Gaussian matrix $N$;
compute $R = NA$;
compute $Y = R(A^{-1} - G)$;
perform an ID of $Y$ to precision $\varepsilon$:

$$\|Y - Y^{(skel)}P\| \leq \varepsilon$$

let $\{j_1, \ldots, j_k\}$ be indices of the $k$ columns of $Y$ that form $Y^{(skel)}$;
solve linear systems:

$$Au_i = (I - AG)(\cdot, j_i), \quad i = 1, \ldots, k$$

and form $B = [u_1|\cdots|u_k]$;
set $\delta G = BP$;

construct the low-rank correction by selecting values for $l$ and $\varepsilon$ and following
the procedure described in Algorithm 1. At this preliminary stage, we measure
the effectiveness of the preconditioner in terms of GMRES iterations needed
to reach a set error tolerance, and regard the banded Jacobi preconditioner as
a reference for the impact of the low-rank correction.

We investigate the behavior of the preconditioner for (uniform) orders of approxi-
mation $p = 2, \ldots, 8$, set $l$ to be half of the matrix dimension, and decrease
$\varepsilon$ in order to obtain approximations of growing rank. The results are recast
in terms of (average) number of GMRES iterations per relative rank, namely
the rank divided by the matrix dimension, see Figure 2 for results relative to
approximations of even order. Although the banded Jacobi preconditioner
is quite effective in the case of symmetric positive definite (SPD) matrices that
arise from discretizations of elliptic problems, the low-rank update does indeed
improve its performance in the case of high orders of approximation. In the
case of lower order approximations, as the rank grows, we observe the desirable
dercrease in the number of GMRES iterations only after an initial, and possibly
substantial, increase. Nevertheless, numerical evidence as presented in Figure
2 suggests that the behavior of the preconditioner improves as the order of
approximation increases. Since we are ultimately interested in developing a
preconditioner for the Helmholtz problem, where high orders of approximation
are crucial, the preconditioner appears to have the desirable properties.

In the case of an SPD problem, it is natural to construct a preconditioner
that is SPD as well. It is possible to recover an SPD preconditioner, by complementing Algorithm 1 with the following steps. Let us replace $\delta G$ by its symmetric part $\delta G^{(\text{sym})}$ and assume the spectral decomposition

$$
\delta G^{(\text{sym})} = V \text{diag} \{ \lambda_1, \ldots, \lambda_n \} V'
$$

where $V = [v_1 | \cdots | v_n]$ is an orthonormal basis of eigenvectors and $\{ \lambda_1, \ldots, \lambda_n \}$ are the eigenvalues ordered from smallest to largest. Let the first $k$ eigenvalues be non-positive and seek an approximation of the form:

$$
(I - V_k \text{diag} \{ c_1, \ldots, c_k \} V_k') \delta G^{(\text{sym})}
$$

where $V_k = [v_1 | \cdots | v_k]$, and $c_1, \ldots, c_k$ are suitable constants to be determined. We immediately see that:

\[
(I - V_k \text{diag} \{ c_1, \ldots, c_k \} V_k') V \text{diag} \{ \lambda_1, \ldots, \lambda_n \} V' = \\
= V \text{diag} \{ \lambda_1, \ldots, \lambda_n \} V' - V_k \text{diag} \{ c_1, \ldots, c_k \} \text{diag} \{ \lambda_1, \ldots, \lambda_k \} V_k' = \\
= V \text{diag} \{ \lambda_1 (1 - c_1), \ldots, \lambda_k (1 - c_k), \lambda_{k+1}, \ldots, \lambda_n \} V'
\]

Thus, the approximation is positive-definite for any choice of constants $c_i$'s such that $\lambda_i (1 - c_i) > 0$ for $i = 1, \ldots, k$. Since the sum of SPD matrices is itself SPD, the preconditioner is SPD provided that $G$ is SPD (which is usually the case.)
Let us turn our attention to an FE discretization of the Helmholtz equation
\[-\Delta u - \kappa^2 u = f\] on \(\Omega\), complemented with a homogenous Neumann boundary condition on \(\partial\Omega\). We employ the same FE discretization as in the case of the Poisson problem, and set the load vector equal to a vector with unit entries. We select non-dimensional wave numbers \(\kappa = 10, 50\) and perform simulations for \(p = 2, \ldots, 8\), see Figure 3 for selected results. As expected, and contrary to the case of the Laplace operator, the low-rank correction is crucial for the convergence of the iterative solver. Every curve features an initial plateau, which indicates that GMRES fails to converge for small values of the rank of the ID. It follows a sharp drop, especially evident in the case \(\kappa = 50\), which conveys that the preconditioner becomes effective once a target rank has been reached. Similarly to the case of the Laplace operator, as the order of approximation increases, the target rank for the low-rank correction to become effective decreases.
4 Application to Topology Optimization

The construction of the preconditioner is tantamount to the solution of \( k \) linear systems. Consequently, it is of practical interest for problems whose nature allows the solution cost to outweigh the construction cost. A favorable scenario arises when the solution of a large number of instances of the same problem, under slowly varying conditions, is needed. In such cases, we can recompute \( G \) at each instance of the varying conditions, while attempting to recycle \( \delta G \) between different instances. Topology optimization fits this framework very well.

Topology optimization is an iterative method, used mainly for PDE constrained optimization problems, to create highly optimized designs by determining a distribution of material that fulfills a specific task in a locally optimal manner without the need to enforce any a priori restrictions on the designs topology, see [1]. Applying the efficient technique of adjoint sensitivity analysis to obtain gradient information for the objective function allows the use of continuous optimization techniques in topology optimization. In this work the Method of Moving Asymptotes, introduced by Svanberg in [16], has been used as the optimization algorithm. The use of gradient-based techniques results in a significant reduction in the number of iterations needed to obtain a good design, compared to other methods. However, for many problems the method still requires \( O(100)-O(1000) \) iterations to recover a locally optimal and physically admissible design. This unavoidably implies that the governing equations for the problem under consideration must be solved a large number of times for a slowly changing material distribution. When PDE problems of several millions of degrees of freedom are considered, as is often the case for real world problems, their solution through traditional direct techniques becomes infeasible. Considering such problems therefore naturally raises the interest in using highly scalable parallel iterative techniques. For physical problems governed by the Helmholtz equation, like acoustic, electromagnetic and structural vibration problems, no general scalable parallel iterative techniques currently exists. Therefore it is of interest to investigate the preconditioner presented in this paper in the context of topology optimization.

We investigate two model problems governed by the Helmholtz equation in order to investigate the preconditioners performance for both reflecting and absorbing boundary conditions. We consider topology optimization of an acoustic cavity and of an acoustic lens constructed from a periodic array of unit cells. We base our formulation of the topology optimization problem on the work done by [4] and the approach presented in [3]. It is emphasized that the considered optimization problems are highly non-convex and therefore small
changes in any parameter value may result in different final designs.

4.1 Pressure Minimization in Acoustic Cavity

We consider the case of an acoustic cavity optimization problem in the domain \( \Omega \subset \mathbb{R}^2 \). The objective is to minimize the average sound pressure in a small target sub-domain of \( \Omega_{\text{OP}} \subset \Omega \), by introducing material acting as a hard wall, in another sub-domain \( \Omega_{\text{d}} \subset \Omega \). The boundary of the cavity \( \delta \Omega \) is taken to be perfectly reflecting, except for a small section \( P \) where a pure tone is excited, see Figure 4. The pressure field \( p \) is governed by the Helmholtz equation with appropriate boundary conditions:

\[
\text{div} \left( \frac{1}{\rho} \nabla p \right) + \frac{\omega^2}{\kappa} p = 0 \quad \text{in } \Omega \quad (1a)
\]
\[
\frac{1}{\rho} \nabla p \cdot \mathbf{n} = 0 \quad \text{on } \text{int}(\partial \Omega \setminus P) \quad (1b)
\]
\[
\frac{1}{\rho} \nabla p \cdot \mathbf{n} = -i\omega U \quad \text{on } P \quad (1c)
\]

Here ‘i’ denotes the imaginary unit, \( \omega = 2\pi f \) is the angular frequency, \( f \) is the frequency, and \( U \) is the vibrational velocity of the source. Like the pressure, the density \( \rho \) and bulk modulus \( \kappa \) depend on the spatial position. The physics of the problem dictates that a given spatial position either contains solid or void. The parameters for the void regions are taken to be those of air at 0% humidity, a temperature of 20°C and a background pressure of 1 atm: \( \rho_{\text{air}} = 1.204 \text{ kg m}^{-3}, \kappa_{\text{air}} = 141.921 \cdot 10^3 \text{ Pa} \). The parameters for the solid are taken to be those of aluminum: \( \rho_{\text{solid}} = 2643 \text{ kg m}^{-3}, \kappa_{\text{solid}} = 6.87 \cdot 10^{10} \text{ Pa} \). This choice for the solid assures that any point containing solid will act as a hard wall, i.e., a zero Neumann boundary condition. An excitation frequency \( f = 90 \text{ Hz} \) and a vibrational velocity \( U = 0.01 \text{ m/s} \) is used. This choice of frequency corresponds to a normalized non-dimensional wavenumber of \( k \approx 30 \).

For the optimization a rescaling of the modeling equation is performed where \( \tilde{\omega} = \frac{\omega}{c}, \tilde{c} = \sqrt{\frac{\rho_{\text{air}}}{\kappa_{\text{air}}}} \), \( \tilde{\rho}_{\text{air}} = 1, \tilde{\rho}_{\text{solid}} = \frac{\rho_{\text{solid}}}{\rho_{\text{air}}}, \tilde{\kappa}_{\text{air}} = 1, \tilde{\kappa}_{\text{solid}} = \frac{\kappa_{\text{solid}}}{\kappa_{\text{air}}} \) is introduced.

In order to employ a continuous optimization approach, an auxiliary field, \( \xi \), such that \( 0 \leq \xi(\mathbf{x}) \leq 1 \) when \( \mathbf{x} \in \Omega_{\text{d}} \), and \( \xi(\mathbf{x}) = 0 \) when \( \mathbf{x} \in \Omega \setminus \Omega_{\text{d}} \), is introduced to interpolate between the inverse material parameters of solid and air. Thus, a location where \( \xi = 1 \) consists of solid material, while a location where \( \xi = 0 \) is occupied by air. A straightforward application of this strategy often results in fragmented and not physically admissible designs, consisting of large areas of intermediate values of \( \xi \). To assure a physically admissible design consisting of solid and air regions only and to control the spatial variations of
the auxiliary field, we employ smoothing and projection. Such operators are defined, respectively, as:

\[
\tilde{\xi}(\xi, x) = \frac{\int_\Omega w(y - x)\xi(y)dy}{\int_\Omega w(y - x)dy}, \quad w(x) = \begin{cases} R - |x|, & |x| < R \\ 0 & \text{otherwise} \end{cases}
\]

\[
\xi(\xi, x) = \frac{\tanh(\beta \eta) + \tanh(\beta(\tilde{\xi}(\xi, x) - \eta))}{\tanh(\beta \eta) + \tanh(\beta(1 - \eta))}.
\]

Here \( R \) is the smoothing radius, \( \eta \) is the projection level, and \( \beta \) is the projection strength. A continuation scheme, where the projection strength is gradually increased, is applied to enforce pure solid and void for the final optimized design while allowing the design to form freely at the beginning of the optimization, see [10, 18]. The interpolation between solid and air is performed as \( \rho^{-1}\xi = \rho_{\text{air}}^{-1} + \xi (\tilde{\rho}_{\text{solid}} - \rho_{\text{air}}^{-1}) \) and \( \kappa^{-1}\xi = \kappa_{\text{air}}^{-1} + \xi (\tilde{\kappa}_{\text{solid}} - \kappa_{\text{air}}^{-1}) \).

For the cavity optimization \( R = 0.72 \) m and \( \eta = 0.5 \) are used. The projection strength \( \beta \) is initiated at 1 and incremented for the continuation scheme as \( \beta_{\text{new}} = 1.2\beta_{\text{old}} \). Its maximal value \( \beta_{\text{max}} \) is set to 100. The increase in \( \beta \) is applied every fifth iteration until \( \beta > 3 \), after which \( \beta \) is increased every fifteenth iteration until \( \beta_{\text{max}} \) is reached.

The minimization of the average of \( p \) in \( \Omega_{\text{OP}} \) may be stated as the continuous optimization problem

\[
\min_{\xi} \Phi(\xi) := \int_{\Omega_{\text{OP}}} |p(x, \xi(x))|^2
\]

where \( p \) is obtained by solving (1) for a given realization of \( \xi \). Adjoint sensitivity analysis is applied to obtain the gradient of the objective function with respect to \( \xi \). The optimization problem is solved with the change in the design variables restricted to \( \Delta \xi = 0.1 \) for each iteration.

The model problem is discretized using a structured quadrilateral mesh and linear finite elements. The auxiliary field \( \xi \) is discretized in a piecewise constant fashion so that each finite element is associated to a single variable determining the value \( \xi \) in that element. These variables are termed “design variables.” The discretization approach, filtering strategy, adjoint sensitivity analysis results, and the continuation approach, see [3], are used in the optimization process.

For the discretization we employ 5 different grids, resulting in approximations of size \( n \in \{8646, 10011, 20301, 31626, 42486\} \), see Figure 5 for a sample matrix. For our test of the preconditioner we limit ourselves to the first 115 iterations of the optimization process. Figure 6 and Figure 7 illustrate the evolution of the design and the pressure field for the case \( n = 31626 \) by showing selected iterations.
We employ the inverse of the diagonal of $A$ as the Jacobi preconditioner $G$. The low-rank correction $\delta G$ is constructed by selecting $l = n/2$ as the oversampling parameter. Numerical experiments, see Figure 9, strongly suggest that, for set $\varepsilon$, the rank of the ID grows linearly with respect to the problem size. In other words, to every $\varepsilon$, there is a corresponding relative rank, independent of the problem size. The performance of the preconditioner is investigated for $\varepsilon = 0.48$ and $\varepsilon = 0.53$, which yield relative ranks of roughly 11% and 7%, respectively.

In all numerical experiments, the low-rank correction $\delta G$ computed at the first step is effective throughout the whole optimization process, namely we do not observe a dramatic increase in the number of GMRES iterations. For example, for $n = 8646$ and $\varepsilon = 0.53$, the initial number of GMRES iterations is 37, while the final number is 57; the maximum number iterations is 74, and it occurs at step 77.

Numerical evidence supports the fact that, as the relative rank of $\delta G$ increases, the total number of GMRES iterations needed to solve the optimization process decreases, see Figure 8. Nevertheless, since the rank of the ID grows linearly with respect to the problem size, the cost of applying the preconditioner is still proportional to $n^2$. The key observation is that the number of GMRES iterations decreases as the problem size increases. In other words, the preconditioner becomes increasingly effective as the size of the problem grows. In fact, the smallest number of iterations is attained for the largest problem size, see Figure 8. Thus, we expect the solution cost to grow slower than $O(n^2)$. This is indeed supported by the numerical experiments, see Figure 9. We clearly see that the solution cost grows slower than $O(n^2)$, and appears to approach $O(n^{1.5})$ asymptotically.

### 4.2 Sound Focusing by Periodic Lens

As a second topology optimization problem, we consider the design of a periodic lens. We consider the problem in $\mathbb{R}^2$, illustrated in Figure 10. As for the previous case, the physics is governed by the Helmholtz equation, complemented with appropriate boundary conditions, see (5). The lens is tasked with focusing the acoustic pressure into a small target area $\Omega_{OP}$, and is constructed as a periodic array of $5 \times 1$ cells, $\Omega_p$, placed in the design domain $\Omega_d$. Axial symmetry is enforced in each cell. The sound pressure is generated by a piston-like source at $P$, placed in a rigid baffle modeled by zero Neumann conditions on $\delta \Omega_{ref}$. For this problem we consider first order absorbing boundary conditions on $\delta \Omega_{abs}$ to model the far field truncation. The resulting boundary
value problem is:
\[
\text{div}\left(\frac{1}{\rho} \nabla p\right) + \frac{\omega^2}{\kappa} p = 0 \quad \text{in } \Omega \quad (5a)
\]
\[
\frac{1}{\rho} \nabla p \cdot \mathbf{n} = 0 \quad \text{on } \partial \Omega_{\text{ref}} \quad (5b)
\]
\[
\nabla p \cdot \mathbf{n} = -i\frac{\omega}{c} p \quad \text{on } \partial \Omega_{\text{abs}} \quad (5c)
\]
\[
\frac{1}{\rho} \nabla p \cdot \mathbf{n} = -i\omega U \quad \text{on } P \quad (5d)
\]

We choose the same material parameter values for the void and solid regions as in the previous case and perform the same rescaling of parameters for the optimization. The excitation frequency and the vibrational velocity of the source are set to \( f = 12 \text{ kHz} \) and \( U = 0.01 \text{ m/s} \) respectively. This choice of frequency corresponds to a normalized non-dimensional wavenumber of \( k \approx 43 \).

We again introduce an auxiliary field, \( \xi \), defined as in the previous case with the additional periodicity condition \( \xi(x) = \xi(x + (2, 0)) \) for all \( x \in \Omega_d \). This is dictated by the fact that we consider a periodic array of cells within \( \Omega \) with a period of 2 in the \( x \)-direction. The periodicity of the array and the axial symmetry about the center of each unit cell substantially reduce the number of design variables. To obtain a physically admissible design, we again employ the smoothing and projection operators, and the continuation scheme that were all described in the previous section. Due to the filtering operators reaching outside the design domain, the periodicity of the design variables does not translate directly to a perfect periodicity of the final design. The filter range is chosen as \( R = 1.25 \text{ cm} \) and the values for \( \eta, \beta, \beta_{\text{max}} \) are set as for the cavity optimization problem. The increment for the continuation scheme is set as: \( \beta_{\text{new}} = 2\beta_{\text{old}} \) and \( \beta \) is increased every 50 iterations until \( \beta_{\text{max}} \) is reached.

The optimization problem is defined formally as in (4), with the additional periodicity requirement on the field \( \xi \). This time the changes in the design variables are restricted to \( \Delta \xi = 0.01 \) for each iteration, and the suggested move limits are used. The same filtering and continuation approach as for the previous example are used. As a consequence of the absorbing boundary conditions, the finite element approximation produces a symmetric, complex-valued stiffness matrix \( A \). Following a standard procedure, we recast the problem into an antisymmetric real problem for the stiffness matrix \( \begin{bmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{bmatrix} \), see Figure 11.

We consider a discretization such that the periodic design cells are discretized using \([80 \times 8]\) finite elements. This choice leads to a total number of degrees of freedom for the discretized Helmholtz problem of \( n = 13122 \) in
the complex system matrix. Again we consider the first 115 iterations of the optimization when investigating the performance of the preconditioner.

Figure 12 and Figure 13 illustrate the evolution of the design and the pressure field by showing selected iterations in the optimization.

As before, we employed the inverse of the diagonal as the Jacobi preconditioner $G$, and selected $l = n/2$ as an oversampling parameter for the construction of $\delta G$. We conducted numerical experiments for ID tolerances $\varepsilon = 0.32$, $\varepsilon = 0.35$ and $\varepsilon = 0.40$, which yields relative ranks of approximately 20\%, 17\% and 15\% respectively. The behavior is qualitatively the same for all cases, see Figure 14. The low-rank correction $\delta G$ computed at the first step is remarkably effective throughout the first half of the optimization process, in the sense that the number of GMRES iterations is nearly constant. Afterwards, although we do observe an increase in the number of iterations, the preconditioner is still effective. In principle, a strategy should be devised to determine whether, at each step of the optimization process, $\delta G$ should be recycled from the previous step or recomputed.

5 Conclusions

We present a general purpose preconditioner built as a low-rank correction to a Jacobi or block-Jacobi preconditioner. The motivation for the construction arises from the exponential decay of the Green function in the context of 2D elliptic problems. This property is routinely exploited in boundary element methods, since it is naturally embedded in the problem discretization. Lately, the properties of the Green function have been the foundation of the so-called sweeping preconditioner. It was originally developed in the context of finite difference approximations of the acoustic wave equation, see [6, 5]. More recently, it was extended to spectral finite elements discretizations of the elastic wave equation, see [17].

In this work we apply related ideas to $hp$-finite elements approximations. Since the inverse stiffness matrix can be understood as a discretization of the Green function, its off-diagonal part, in the case of 2D elliptic problem, is rank-deficient. This allows for building a preconditioner of the form $M = G + \delta G$, where $G$ is a Jacobi or block-Jacobi preconditioner, and $\delta G$ is constructed from the off-diagonal part of the inverse stiffness matrix, and acts as a low-rank correction to $G$. Since $\delta G$ is a low-rank update, $M$ can be applied cheaply.

While elliptic problems provide the motivation to our approach, we are ultimately interested in the performance of the preconditioner in the context of wave propagation problems. As discussed in Section 1, the construction of a good preconditioner for wave propagation problems is still an open problem.
Our preconditioner proves successful in the case of moderate wave numbers, see Section 3, although a low-rank approximation no longer exists \textit{a priori}. The construction of $\delta G$ is rather expensive. In fact, if $k$ is the (small) rank of $\delta G$, it is $k$ times more expensive than a single linear solve. In order to outweigh the cost of construction, thus making our approach competitive, we tested the preconditioner in the context of topology optimization for acoustics. Topology optimization is an iterative technique where the governing equations, \textit{e.g.}, the Helmholtz equation, need to be solved a number of times, \textit{e.g.}, 1000 times, for a slow changing design. We computed $\delta G$ at the first iteration and were able to employ it throughout the entire optimization process, see Section 4.

Since we are ultimately interested in solving large systems of equations, \textit{i.e.}, several millions of unknowns, scaling properties are crucial. Although we have no theoretical explanation, numerical evidence suggests that solution cost is $O(n^{1.5})$. Furthermore, this result should be dimension-independent and thus carry over to 3D problems.

As future directions of research, we shall address the parallelization of the preconditioner implementation.

References


Figure 3: Discrete Helmholtz problem. Average number of GMRES iterations as a function of the ID rank for different orders of approximation. The rank is reported as relative to the matrix size. Two different wave numbers, $\kappa = 10$ and $\kappa = 50$, were considered.
Figure 4: Acoustic cavity problem diagram. All measurements are given in meters. Domain: $\Omega = [0, 18] \times [0, 9]$. Target sub-domain: $\Omega_{OP} = [6, 18] \times [0, 2]$. Design sub-domain: $\Omega_d = [0, 18] \times [6, 9]$. Reflecting boundary: $\delta\Omega$. Excitation boundary: $P = 0 \times [1.3, 1.7]$.

Figure 5: Acoustic cavity problem stiffness matrix. The problem discretization gives rise to a nine-diagonal real-valued matrix.
Figure 6: Snapshots of design evolution across the 115 optimization iterations. The coordinate axes shows the elements in the finite element discretization. White corresponds to air and black corresponds to solid.
Figure 7: Snapshots of pressure field evolution, across the 115 optimization iterations. The pressure is reported in dB, $\text{SPL}_{\text{ref}=20\mu \text{Pa}}$. The coordinate axes shows the nodes in the finite element discretization.
Figure 8: Cumulative number of GMRES iterations needed to solve the first 10 iterations of the acoustic cavity topology optimization problem for different problem sizes $n$. Solid lines refer to results obtained with $\varepsilon = 0.53$, dashed lines to results obtained with $\varepsilon = 0.48$. Matching markers refer to the same problem size $n$. 
Figure 9: Dependence of the ID rank and solution time upon the problem size. Circles refer to results obtained for $\varepsilon = 0.53$, diamonds refer to results obtained for $\varepsilon = 0.48$. 
Figure 10: Acoustic lens problem diagram. All measurements are given in centimeters. Domain: $\Omega = [0, 20] \times [0, 20]$. Target sub-domain: $\Omega_{OP} = [16.4, 19.6] \times [9.4, 10.6]$. Design sub-domain: $\Omega_d = [4, 14] \times [0, 20]$. Periodic design cell: $\Omega_p = [4 + 2n, 6 + 2n] \times [0, 20], n \in \{0, 1, 2, 3, 4\}$. Absorbing boundary: $\delta \Omega_{abs}$. Reflecting boundary: $\delta \Omega_{ref}$. Excitation boundary condition: $P = 0 \times [9.8, 10.2]$. 
Figure 11: Acoustic lens problem stiffness matrix. As a consequence of the absorbing boundary conditions, the finite elements stiffness matrix $A$ (left) is symmetric and complex-valued. By separating the real and imaginary part, we obtain the antisymmetric real-valued matrix $\begin{bmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{bmatrix}$ (right).
Figure 12: Snapshots of design evolution across the optimization. The coordinate axes shows the elements in the finite element discretization. Black corresponds to solid and white corresponds to air.
Figure 13: Snapshots of pressure field evolution, across the optimization. The pressure is reported in dB, \( \text{SPL}_{\text{ref}=20\mu \text{Pa}} \). The coordinate axes shows the nodes in the finite element discretization.
Figure 14: Number of GMRES iterations needed to solve the acoustic lens topology optimization problem. The ID tolerances $\varepsilon = 0.30, 0.32, 0.35$ correspond to a relative ranks of about 20\%, 17\% and 15\% respectively. Although a marked increase in the number of GMRES iterations is observed in the second half of the considered part of the optimization process, the preconditioner appears to be remarkably effective.